Bridging Cognitive and Quantitative Psychology

Lesa Hoffman

Associate Professor and Director
Cognitive and Quantitative Program
Department of Psychology
University of Nebraska-Lincoln

Presented 10/11/12 at the annual meeting of the
***Society for Multivariate Experimental Psychology,* Vancouver, BC.**

Once Upon a Time…

October 27, 2005

Dear Quantitative Search Committee Members:

I was pleased to encounter the advertisement for the position of tenure-track Assistant Professor in Psychology at the University of Nebraska – Lincoln, particularly because of your request for an emphasis in quantitative methodology. I believe that my unique quantitative strengths in multilevel, structural, and item response modeling, coupled with my independent and collaborative research experiences within cognitive and developmental psychology, would complement both the research activities and the training goals of your department. I remember fondly my days at UNL as an undergraduate psychology major, and I am particularly excited by the possibility of contributing at a new level to the program and faculty that were so integral in shaping my early professional development….

Burnett Hall, home of UNL Psych

Breakfast area at Embassy Suites hotel

Rick Bevins

Data analyses for acquisition. SAS PROC MIXED or NLMIXED (for any nonlinear parameters) will be used to estimate all models. First, to examine the acquisition data, we will use 2 versions of multivariate multilevel model in which sessions are treated as nested within rats, as illustrated in Equations 1 and 2:

$$\begin{aligned}
&\text{Level 1: } \text{Dipper}_{tid} = (\text{Sal})[\beta_{0is} + \beta_{1is}(\text{Session}_{tis}) + \beta_{2is}(\text{Session}_{tis})^2 + \beta_{(3\text{-}5)is}(\text{Estrous}_{tis}) + e_{tis}] + \\
&\qquad (\text{Nic})[\beta_{0in} + \beta_{1in}(\text{Session}_{tin}) + \beta_{2in}(\text{Session}_{tin})^2 + \beta_{(3\text{-}5)in}(\text{Estrous}_{tin}) + e_{tin}] \qquad (1)\\
&\text{Level 2: } \beta_{0is} = \gamma_{00s} + U_{0is} \quad \beta_{1is} = \gamma_{10s} + U_{1is} \quad \beta_{2is} = \gamma_{20s} + U_{2is} \quad \beta_{(3\text{-}5)is} = \gamma_{3\text{-}50s}\\
&\qquad \beta_{0in} = \gamma_{00n} + U_{0in} \quad \beta_{1in} = \gamma_{10n} + U_{1in} \quad \beta_{2in} = \gamma_{20n} + U_{2in} \quad \beta_{(3\text{-}5)in} = \gamma_{3\text{-}50n}
\end{aligned}$$

$$\begin{aligned}
&\text{Level 1: } \text{Dipper}_{tid} = \beta_{0is} + \beta_{1is}(\text{Session}_{tis}) + \beta_{2is}(\text{Session}_{tis})^2 + \beta_{(3\text{-}5)is}(\text{Estrous}_{tis}) + (\text{Sal})e_{tis} + \\
&\qquad (\text{Nic})[\beta_{0in} + \beta_{1in}(\text{Session}_{tin}) + \beta_{2in}(\text{Session}_{tin})^2 + \beta_{(3\text{-}5)in}(\text{Estrous}_{tin})] + (\text{Nic})e_{tin} \qquad (2)\\
&\text{Level 2: } \beta_{0is} = \gamma_{00s} + (\text{Sal})U_{0is} \quad \beta_{1is} = \gamma_{10s} + (\text{Sal})U_{1is} \quad \beta_{2is} = \gamma_{20s} + (\text{Sal})U_{2is} \quad \beta_{(3\text{-}5)is} = \gamma_{(3\text{-}5)0s}\\
&\qquad \beta_{0in} = \gamma_{00n} + (\text{Nic})U_{0in} \quad \beta_{1in} = \gamma_{10n} + (\text{Nic})U_{1in} \quad \beta_{2in} = \gamma_{20n} + (\text{Nic})U_{2in} \quad \beta_{(3\text{-}5)in} = \gamma_{(3\text{-}5)0n}
\end{aligned}$$

In Eq. 1, the level-1 model predicts the dipper entries/sec for each session *t*, rat *i*, and each drug (saline *s* or nicotine *n*). As shown, (Sal) and (Nic) are dummy codes that indicate which kind of session each observation was, such that separate models predicting change across sessions are then specified for each drug. Given the trends in Figure 1, Eq. 1 specifies a quadratic pattern of change across sessions, although other forms (e.g., exponential) will also be evaluated to best describe the overall change across sessions for each drug. The level-1 model in Eq. 1 also includes 3 contrasts ($\beta_{(3\text{-}5)i}$) to distinguish among the 4 estrous phases, as well as a model residual. The level-2 model then describes individual variation across rats in each of the level-1 terms per drug: the intercept (β_{0i}), linear and quadratic rates of change across acquisition sessions (β_{1i} and β_{2i}), and effects of estrous ($\beta_{3\text{-}5i}$). Each is first defined by a fixed effect (γ) representing the sample average effect. The intercept, linear, and quadratic models then also include a random effect (U) representing individual variation across rats. Covariances will be estimated among the random effects, and the significance of fixed and random effects will be evaluated using Wald's tests (estimate/standard error) and likelihood ratio tests, respectively. To summarize, using Eq. 1 we can examine the change across sessions and test estrous differences directly for each drug (nicotine or saline). However, because the presence of acquisition is indicated by *the difference* between the saline and nicotine sessions, we will also use Eq. 2, which is a re-parameterized but equivalent model to Eq. 1. In Eq. 2, however, saline is considered the reference drug, and the model then estimates differences between saline and nicotine in each parameter. Thus, Eq. 2 will tell us how *the difference* in dipper entries/sec between saline and nicotine is moderated by session and estrous phase. By combining results across models, we can determine whether estrous affects dipper entries/sec in saline sessions, in nicotine sessions, and whether it affects the difference between saline and nicotine (reflecting acquisition) at a given session.

Tales from Quognitive Psychology

- Uses (and Misuses) of Multilevel Models
  - Modeling common persons and (un)common items
  - Modeling individual differences in experimental manipulations
  - Modeling extreme repeated measures data

- Combining Experiments and Psychometrics
  - Considerations in blending disparate paradigms
  - Measuring individual differences in selective visual attention
  - Decomposition of item difficulty for creating adaptive tests

- Moving Forward
  - (Non-unique) challenges to overcome
  - Opportunities for research synergy

Studying Cognition using Experiments

- Properties of cognitive abilities (e.g., attention, memory, executive function) are usually examined with experiments involving multi-way repeated measures factorial designs

- e.g., DV = RT, trial manipulations A and B, each with 2 levels:
  - Step 1: Compute mean across trials per unique design condition
  - **Step 2: RM ANOVA on person condition means**

Original Data per Person

	B1	B2
A1	Trial 001 Trial 002 …… Trial 100	Trial 101 Trial102 …… Trial 200
A2	Trial 201 Trial 202 …… Trial 300	Trial 301 Trial302 …… Trial 400

2. Model for condition *c* and person *p* outcome:

$$RT_{cp} = \gamma_0 + \gamma_1 A_c + \gamma_2 B_c + \gamma_3 A_c B_c + U_{0p} + e_{cp}$$

1. →

Summary Data per Person

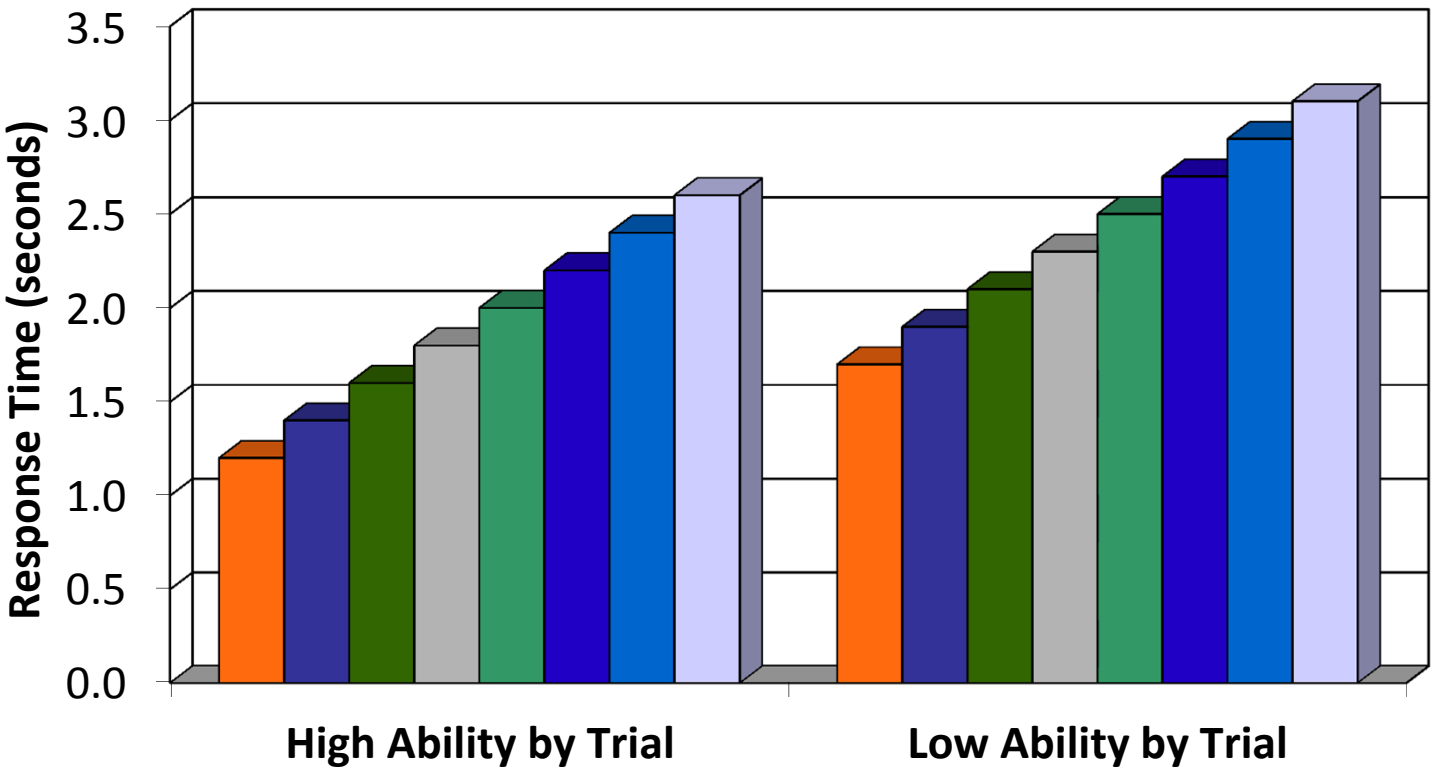
	B1	B2
A1	Mean (A1, B1)	Mean (A1, B2)
A2	Mean (A2, B1)	Mean (A2, B2)

RM ANOVA on Summary Data

- Although complexity for complexity's sake isn't useful…
- May cause significant problems with respect to:
  - Inference about missing responses
  - Discretizing of trial-specific predictors
  - Significance of trial-specific effects may be overly optimistic
  - Effect sizes for trial-specific effects may be overly optimistic
- Limits questions that can be asked about:
  - Adequacy of control of experimental stimuli
  - Moment-to-moment indicators of cognitive processes
  - Individual differences in cognitive processes

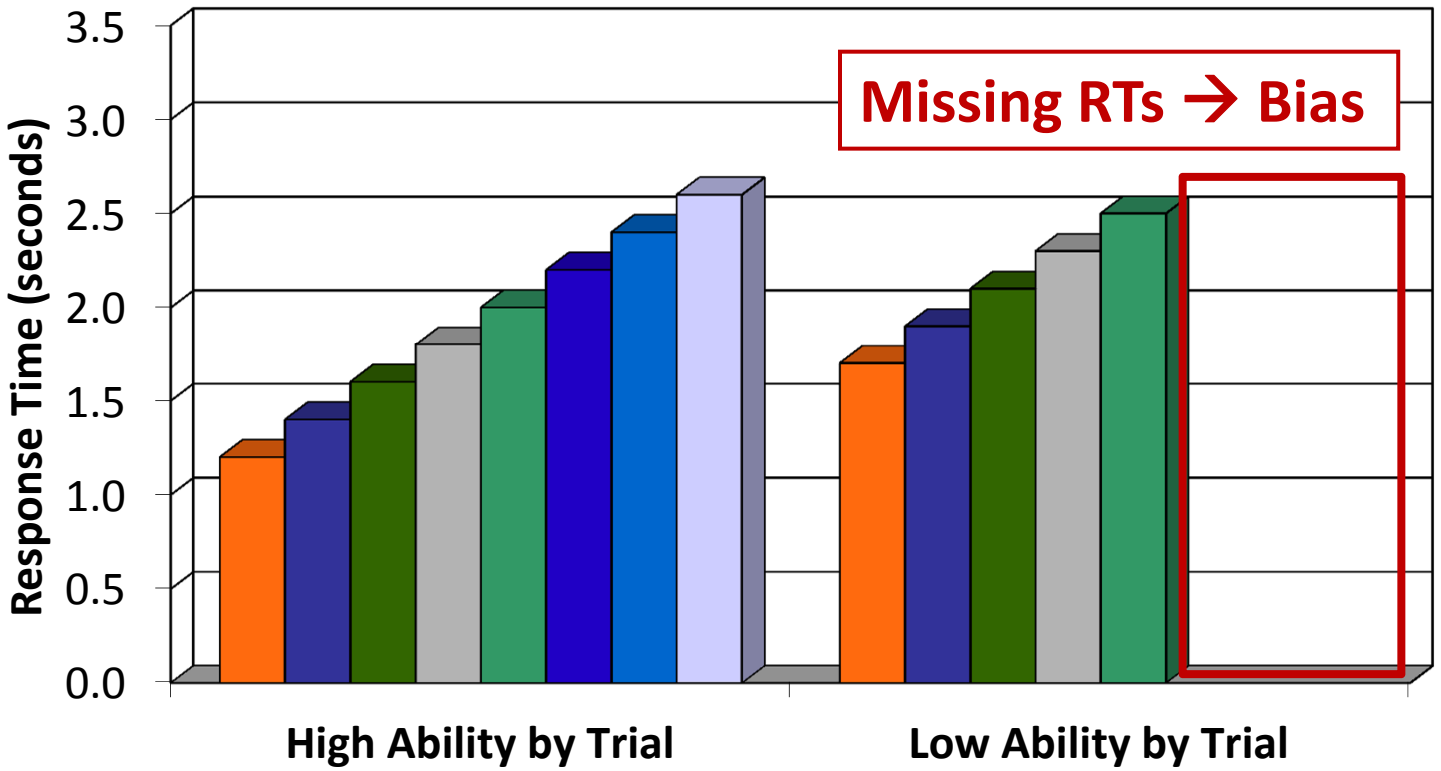
Summary RM ANOVA could be ok when…

- One has ***complete data***
  - e.g., if outcome is response time and accuracy is near ceiling
  - …What if data are not missing *completely* at random (e.g., inaccuracy)?



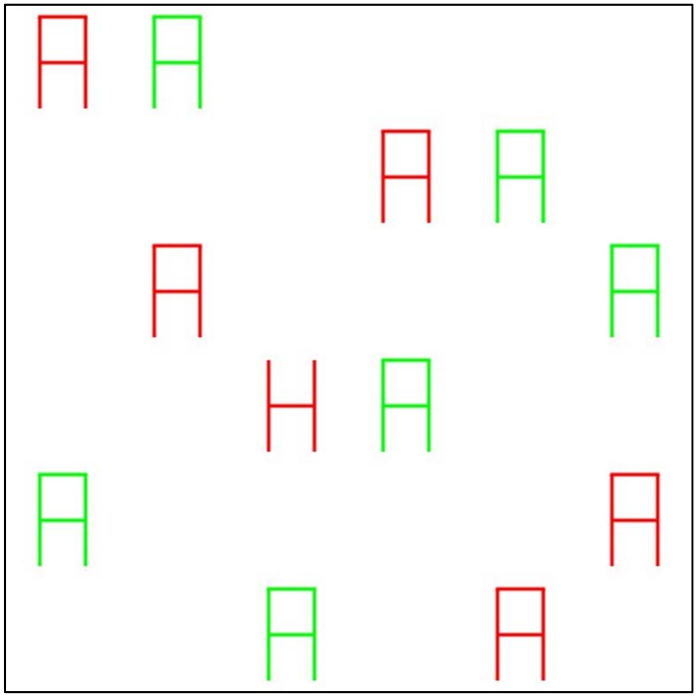
Summary RM ANOVA is *not* ok when…

- Responses are not missing completely at random
  - RTs for inaccurate responses to difficulty trials are not included

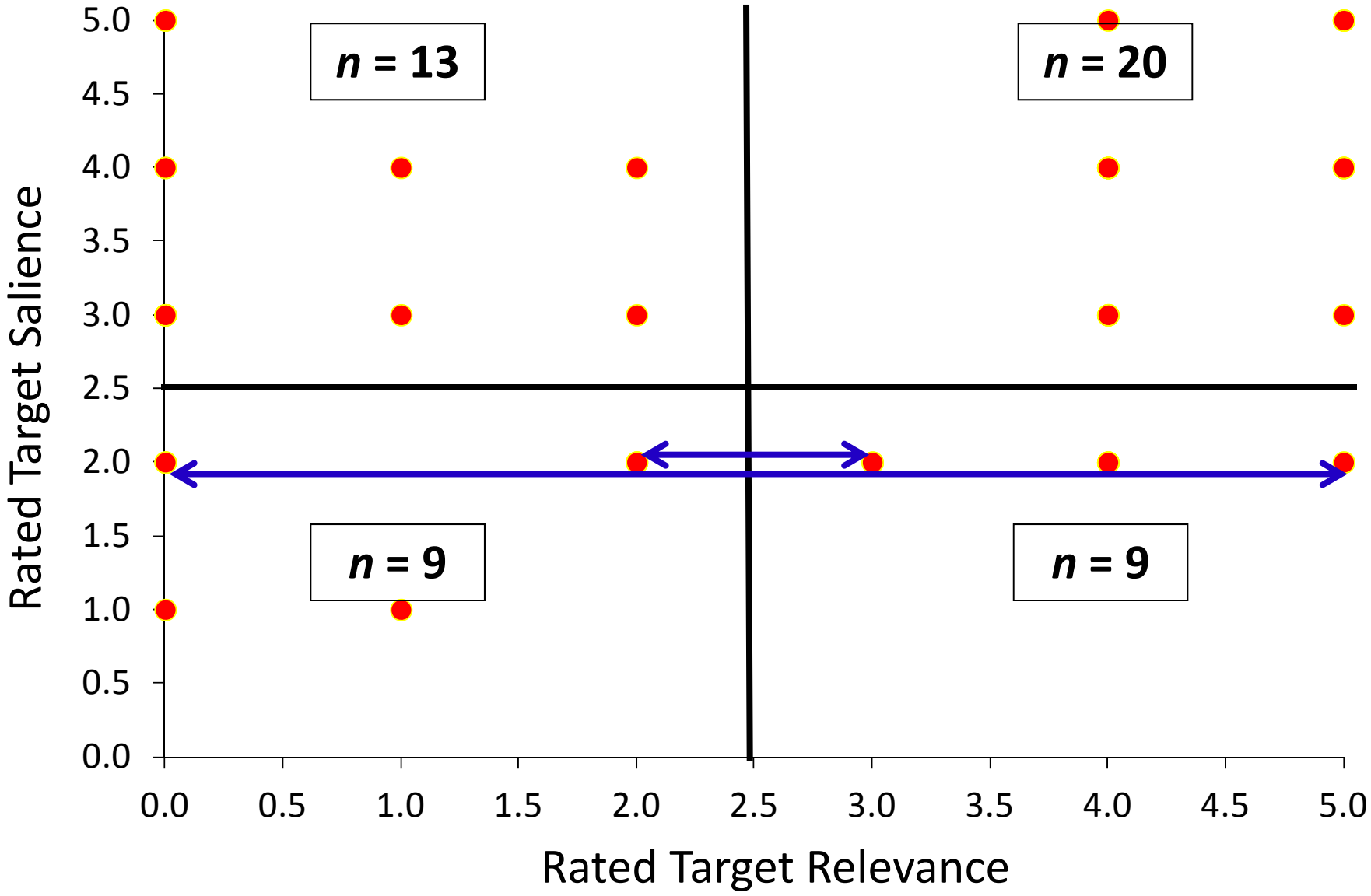


Summary RM ANOVA may be ok when…

- **Trial-specific predictors are truly *discrete conditions***
  - e.g., distinguish a target object by color or by shape
  - e.g., distinguish a target object from 6 vs. 11 non-target objects
  - …What about *continuous* predictors (e.g., target visual salience, relevance)?

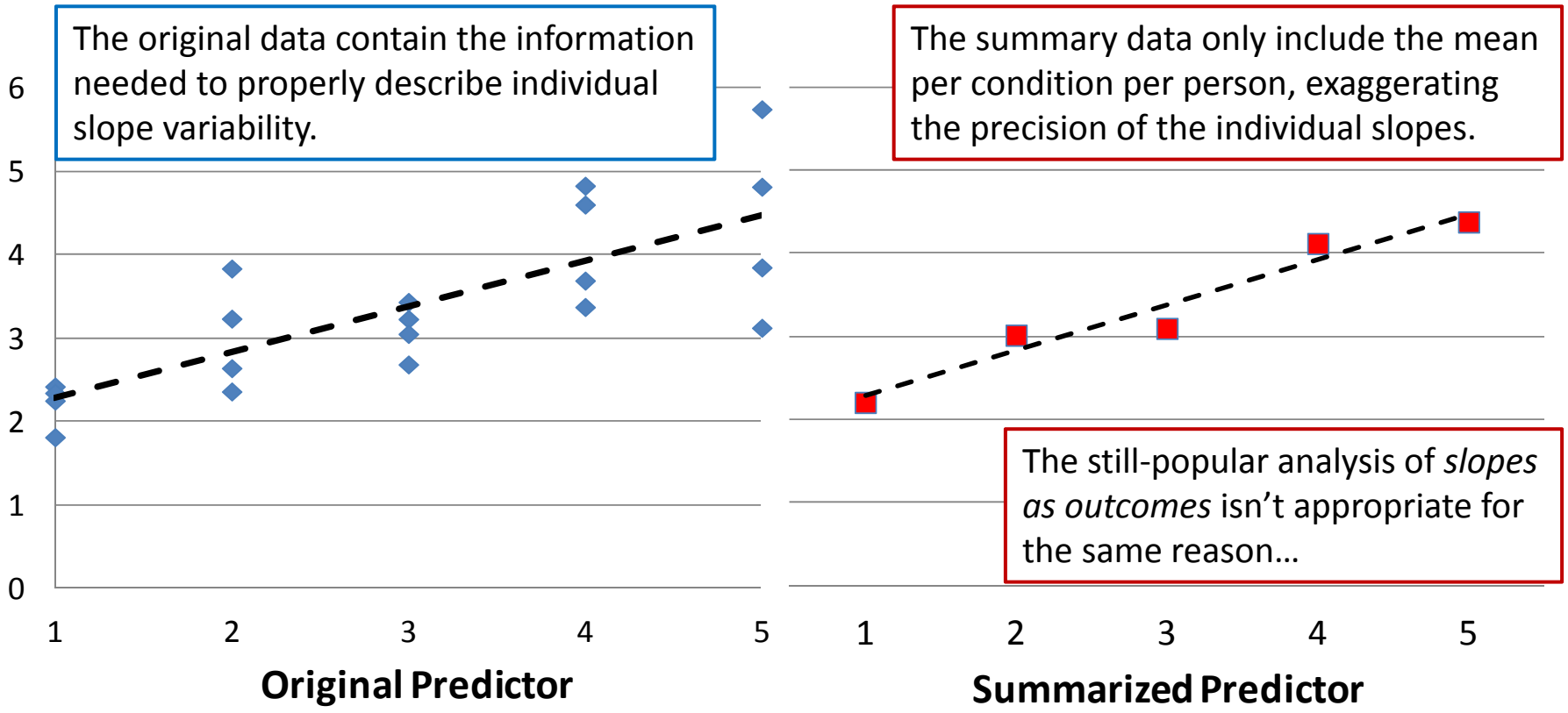


Creating "Conditions" (r = .22 → r ≈ 0)



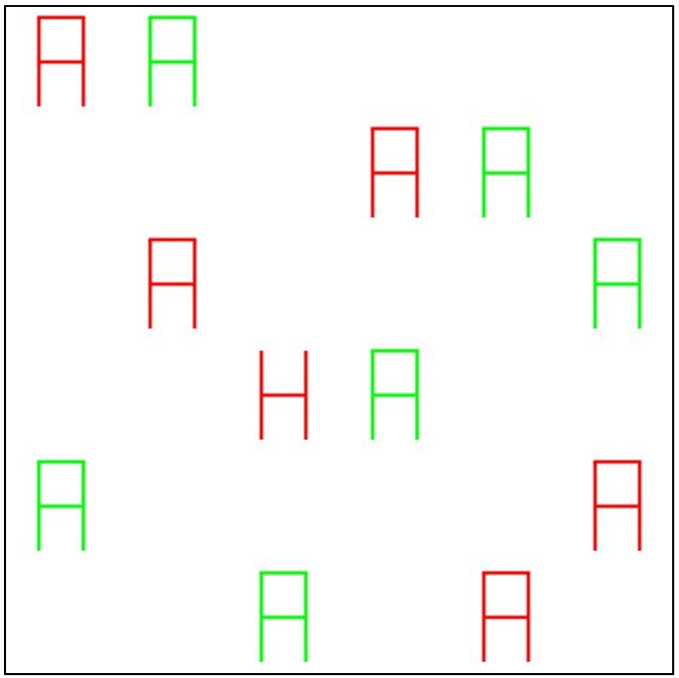
Summary RM ANOVA may be ok when…

- Individual differences expected or observed in means (intercepts) only
- …What about variability in the ***effects of trial-specific manipulations***?



Summary RM ANOVA may be ok when…

- Experimental stimuli are ***controlled and exchangeable***
  - **Controlled** → Constructed, not sampled from a population
  - **Exchangeable** → Stimuli vary only in dimensions of interest
  - …What about *non-exchangeable* stimuli (e.g., real-world scenes)?



The Curse of Non-Exchangeable Items

Jim Bovaird, University of Nebraska-Lincoln

Larry Locker, Georgia Southern University

Downtown Lawrence, KS

- Psycholinguistic research (trials are words and non-words)
  - Common persons, common trials designs
  - Contentious fights with reviewers about adequacy of experimental control when using real words as stimuli
  - Long history of debate as to how words as experimental stimuli should be analyzed… F1 or F2 (or both)?

An Alternative Summary ANOVA

Original Data per Person

	B1	B2
A1	Trial 001 Trial 002 …… Trial 100	Trial 101 Trial102 …… Trial 200
A2	Trial 201 Trial 202 …… Trial 300	Trial 301 Trial302 …… Trial 400

Person Summary Data

	B1	B2
A1	Mean (A1, B1)	Mean (A1, B2)
A2	Mean (A2, B1)	Mean (A2, B2)

"F1" RM ANOVA model on *N* persons:

$$RT_{cp} = \gamma_0 + \gamma_1 A_c + \gamma_2 B_c + \gamma_3 A_c B_c + \boldsymbol{U_{0p}} + e_{cp}$$

"F2" Between-Groups ANOVA model on *T* trials:

$$RT_t = \gamma_0 + \gamma_1 A_t + \gamma_2 B_t + \gamma_3 A_t B_t + e_t$$

Trial Summary Data

	B1
A1, B1	Trial 001 = Mean(Person 1, Person 2,… Person *N*) Trial 002 = Mean(Person 1, Person 2,… Person *N*) …… Trial 100
A1, B2	Trial 101 = Mean(Person 1, Person 2,… Person *N*) Trial 102 = Mean(Person 1, Person 2,… Person *N*) …… Trial 200
A2, B1	Trial 201 = Mean(Person 1, Person 2,… Person *N*) Trial 202 = Mean(Person 1, Person 2,… Person *N*) …… Trial 300
A2, B2	Trial 301 = Mean(Person 1, Person 2,… Person *N*) Trial 302 = Mean(Person 1, Person 2,… Person *N*) …… Trial 400

Choosing Amongst ANOVA Models

- F1 RM ANOVA on **person** summary data:
  - Assumes trials are fixed—within-condition ***trial*** variability is gone
- F2 ANOVA on **trial** summary data:
  - Assumes persons are fixed—within-trial ***person*** variability is gone

- Proposed ANOVA-based resolutions:
  - **F′** → quasi-F test that treats both trials and persons as random (Clark, 1973), but requires complete data (least squares)
  - **Min F′** → lower-bound of F′ derived from F1 and F2 results, which does not require complete data, but is (too) conservative
  - **F1 x F2 criterion** → effects are only "real" if they are significant in **both F1 and F2 models** (aka, death knell for psycholinguists)

  - But neither model is complete (two wrongs don't make a right)…

Multilevel Models: A New Way of Life!

- Level 1: $y_{ij} = \beta_{0i} + \beta_{1j}X1_{ij} + \beta_{2j}X2_{ij} + \cdots + e_{ij}$

- Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}Z1_j + \gamma_{02}Z2_j + \cdots + U_{0j}$

  $\beta_{1j} = \gamma_{10} + \gamma_{11}Z1_j + \gamma_{12}Z2_j + \cdots + U_{1j}$

  $\beta_{2j} = \gamma_{20} + \gamma_{21}Z1_j + \gamma_{22}Z2_j + \cdots + U_{2j}$

- **Levels** are defined within the context of a study, such as:
  - Occasions nested within Persons
  - Persons nested within Groups
  - Groups nested within Locations
  - Items nested within Persons (CFA, IRT)

  - **Everything is nested within *Something*!**

Multilevel Models: A New Way of Life?

Original Data per Person

	B1	B2
A1	Trial 001 Trial 002 …… Trial 100	Trial 101 Trial102 …… Trial 200
A2	Trial 201 Trial 202 …… Trial 300	Trial 301 Trial302 …… Trial 400

Pros:

- Use all original data, not summaries
- Responses can be missing at random
- Can include continuous trial predictors

Cons:

- **Is still wrong**

Level 1: $y_{tp} = \beta_{0p} + \beta_{1p}A_{tp} + \beta_{2p}B_{tp} + \beta_{3p}A_{tp}B_{tp} + e_{tp}$

Level 2: $\beta_{0p} = \gamma_{00} + U_{0p}$

$\beta_{1p} = \gamma_{10}$

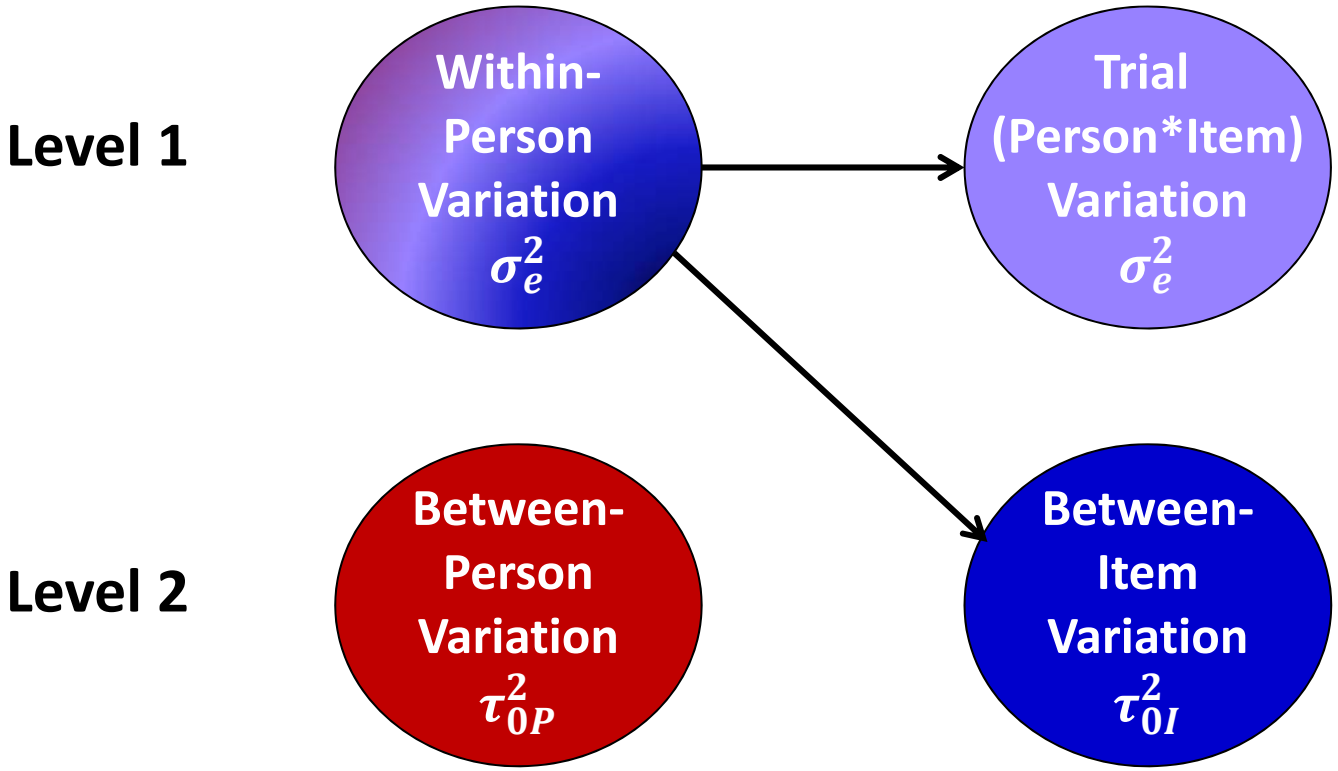
$\beta_{2p} = \gamma_{20}$

$\beta_{3p} = \gamma_{30}$

Level 1 = Within-Person Variation (Across Trials)

Level 2 = Between-Person Variation

Multilevel Models: A New Way of Life?



A Better Way of (Multilevel) Life

Random effects over **persons** of **item** or **trial** predictors can also be tested and predicted.

- **Multilevel Model with *Crossed* Random Effects:**

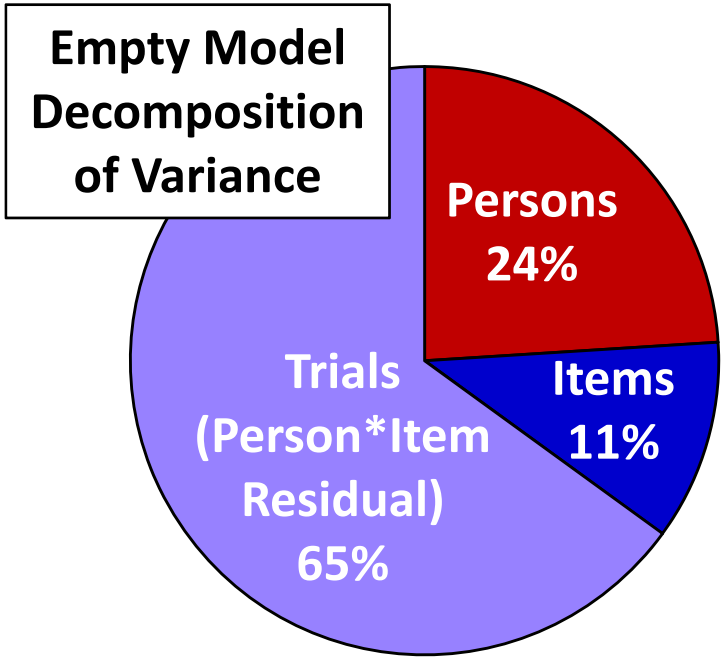
$$RT_{tpi} = \gamma_{000} + \gamma_{001}A_i + \gamma_{002}B_i + \gamma_{003}A_iB_i + \boldsymbol{U_{0p0}} + \boldsymbol{U_{00i}} + \boldsymbol{e_{tpi}}$$

t trial
p person
i item

- Explicitly test **persons** and **items** as random effects:
  - Person predictors capture between-person mean variation: $\boldsymbol{\tau^2_{0P}}$
  - Item predictors capture between-item mean variation: $\boldsymbol{\tau^2_{0I}}$
  - Trial predictors capture trial-specific residual variation: $\boldsymbol{\sigma^2_e}$

Example from Psycholinguistics

- Crossed design: 38 persons by 39 items (words or nonwords)
- Lexical decision task: RT to decide if word or nonword
- 2 word-specific predictors of interest:
  - A: Low/High Phonological Neighborhood Frequency
  - B: Small/Large Semantic Neighborhood Size



Model and Results

$$RT_{tpi} = \gamma_{000} + \gamma_{001}A_i + \gamma_{002}B_i + \gamma_{003}A_iB_i + U_{0p0} + U_{00i} + e_{tpi}$$

Pseudo-R^2:

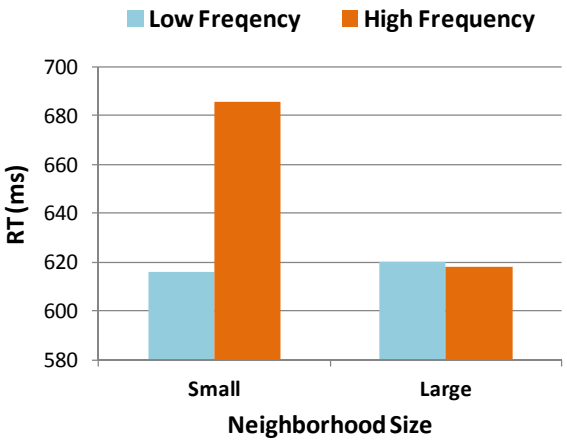
Residual ≈ 0%

Person ≈ 0%

Items ≈ 30%*

Total R^2 ≈ 3.3%

***Significant item variability remained**



Tests of Fixed Effects by Model

	A: Frequency Marginal Main Effect	B: Size Marginal Main Effect	A*B: Interaction of Frequency by Size
F_1 Person ANOVA	F (1,37) = 16.1 p = .0003	F (1,37) = 14.9 p = .0004	F (1,37) = 38.2 p < .0001
F_2 Words ANOVA	F (1,35) = 5.3 p = .0278	F (1,35) = 4.5 p = .0415	F (1,35) = 5.7 p = .0225
F' min (via ANOVA)	F (1,56) = 4.0 p = .0530	F (1,55) = 3.5 p = .0710	F (1,45) = 5.0 p = .0310
Crossed MLM	F (1,32) = 5.4 p = .0272	F (1,32) = 4.6 p = .0393	F (1,32) = 6.0 p = .0199

Tests of Fixed Effects by Model

	A: Frequency Marginal Main Effect	B: Size Marginal Main Effect	A*B: Interaction of Frequency by Size
F_1 Person ANOVA	F (1,37) = 16.1 p = .0003	F (1,37) = 14.9 p = .0004	F (1,37) = 38.2 p < .0001
F_2 Words ANOVA	F (1,35) = 5.3 p = .0278	F (1,35) = 4.5 p = .0415	F (1,35) = 5.7 p = .0225
F' min (via ANOVA)	F (1,56) = 4.0 p = .0530	F (1,55) = 3.5 p = .0710	F (1,45) = 5.0 p = .0310
Crossed MLM	F (1,32) = 5.4 p = .0272	F (1,32) = 4.6 p = .0393	F (1,32) = 6.0 p = .0199

Tests of Fixed Effects by Model

	A: Frequency Marginal Main Effect	B: Size Marginal Main Effect	A*B: Interaction of Frequency by Size
F_1 Person ANOVA	F (1,37) = 16.1 p = .0003	F (1,37) = 14.9 p = .0004	F (1,37) = 38.2 p < .0001
F_2 Words ANOVA	F (1,35) = 5.3 p = .0278	F (1,35) = 4.5 p = .0415	F (1,35) = 5.7 p = .0225
F' min (via ANOVA)	F (1,56) = 4.0 p = .0530	F (1,55) = 3.5 p = .0710	F (1,45) = 5.0 p = .0310
Crossed MLM	F (1,32) = 5.4 p = .0272	F (1,32) = 4.6 p = .0393	F (1,32) = 6.0 p = .0199

Simulation: Type I Error

Condition		Models					
Item Variance	Person Variance	1: Both Random Effects	2: Random Persons Only	3: Random Items Only	4: No Random Effects	5: F1 Person ANOVA	6: F2 Item ANOVA
Item Effect:							
2	2	**0.03**	0.09	0.03	0.09	0.09	0.03
2	10	**0.05**	0.14	0.05	0.12	0.15	0.05
10	2	**0.04**	0.32	0.04	0.31	0.32	0.04
10	10	**0.05**	0.31	0.05	0.29	0.33	0.05
Person Effect:							
2	2	**0.04**	0.04	0.12	0.11	0.04	0.12
2	10	**0.05**	0.05	0.34	0.34	0.05	0.36
10	2	**0.04**	0.03	0.12	0.09	0.03	0.12
10	10	**0.06**	0.06	0.34	0.31	0.05	0.37

Model Items as Fixed → Wrong Item Effect

Condition		Models					
Item Variance	Person Variance	1: Both Random Effects	**2: Random Persons Only**	3: Random Items Only	4: No Random Effects	**5: F1 Person ANOVA**	6: F2 Item ANOVA
Item Effect:							
2	2	0.03	**0.09**	0.03	0.09	**0.09**	0.03
2	10	0.05	**0.14**	0.05	0.12	**0.15**	0.05
10	2	0.04	**0.32**	0.04	0.31	**0.32**	0.04
10	10	0.05	**0.31**	0.05	0.29	**0.33**	0.05
Person Effect:							
2	2	0.04	0.04	0.12	0.11	0.04	0.12
2	10	0.05	0.05	0.34	0.34	0.05	0.36
10	2	0.04	0.03	0.12	0.09	0.03	0.12
10	10	0.06	0.06	0.34	0.31	0.05	0.37

Model Persons as Fixed → Wrong Person Effect

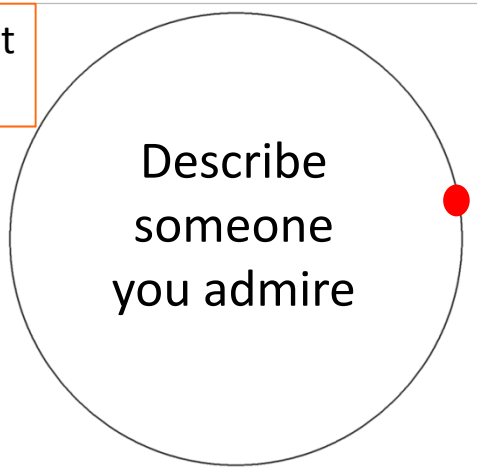
Condition		Models					
Item Variance	Person Variance	1: Both Random Effects	2: Random Persons Only	**3: Random Items Only**	4: No Random Effects	5: F1 Person ANOVA	**6: F2 Item ANOVA**
Item Effect:							
2	2	0.03	0.09	0.03	0.09	0.09	0.03
2	10	0.05	0.14	0.05	0.12	0.15	0.05
10	2	0.04	0.32	0.04	0.31	0.32	0.04
10	10	0.05	0.31	0.05	0.29	0.33	0.05
Person Effect:							
2	**2**	0.04	0.04	**0.12**	0.11	0.04	**0.12**
2	**10**	0.05	0.05	**0.34**	0.34	0.05	**0.36**
10	**2**	0.04	0.03	**0.12**	0.09	0.03	**0.12**
10	**10**	0.06	0.06	**0.34**	0.31	0.05	**0.37**

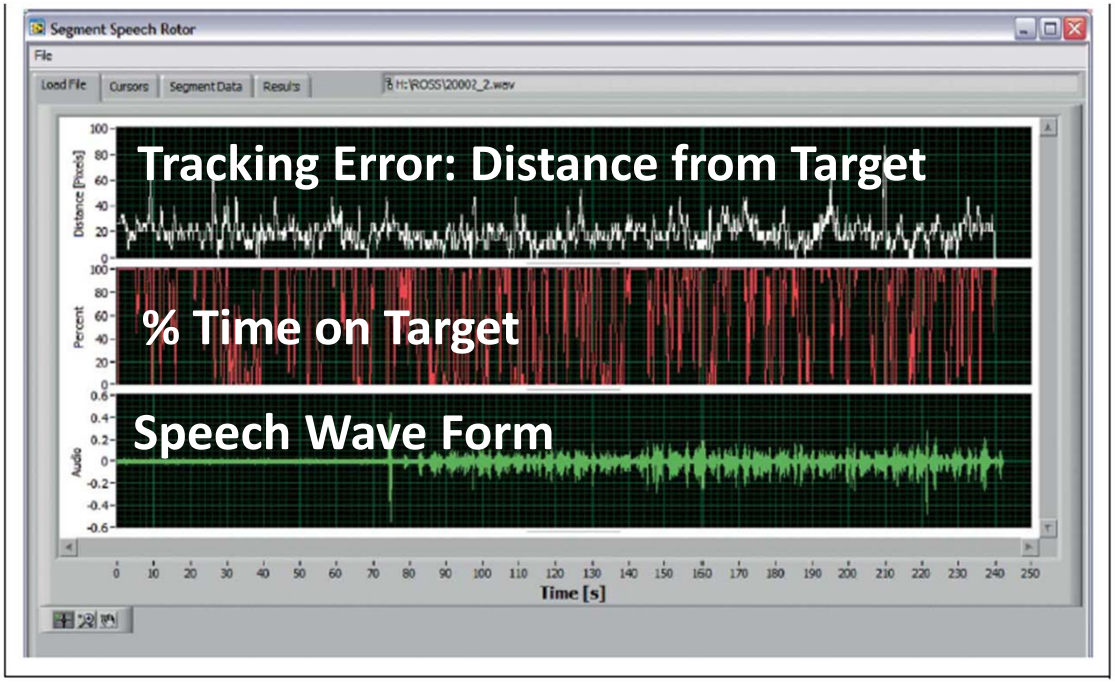
Combining Experimental Designs with MLM

- ANOVAs on **summary data** is problematic:
  - **Ignoring missing** responses; **discretizing** item predictors
  - **Significance and effect sizes** of item-specific effects will be distorted if items are not exchangeable but they are modeled that way
- **MLM** solves these problems and provides a way to quantify and **predict individual differences** in cognitive processes, such as…
  - Forgetting rates over time
    MacDonald, Stigdotter-Neely, Derwinger, & Bäckman (2006, *JEP: General*)
  - Executive function and semantic processing in verbal fluency
    McDowd, Hoffman, Rozek, Lyons, Pahwah, Burns, & Kemper (2011, *Neuropsychology*)
  - Reliance on object-based and space-based deployment of visual attention
    Kliegl, Wei, Dambacher, Yan, & Zhou (2010, *Frontiers in Quantitative Psychology and Measurement*)
- MLM can also be used to examine cognitive processes as they unfold in **extreme repeated measures data**, such as…

Tracking and Talking

Susan Kemper at Fraser Hall, KU

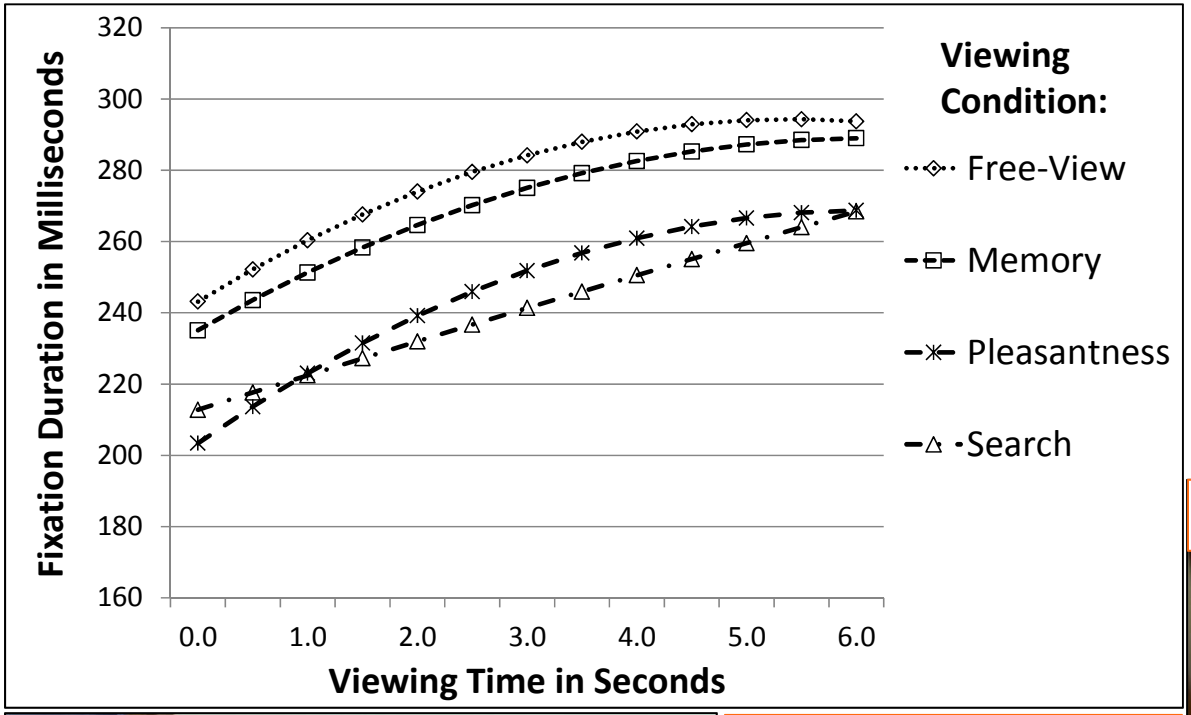


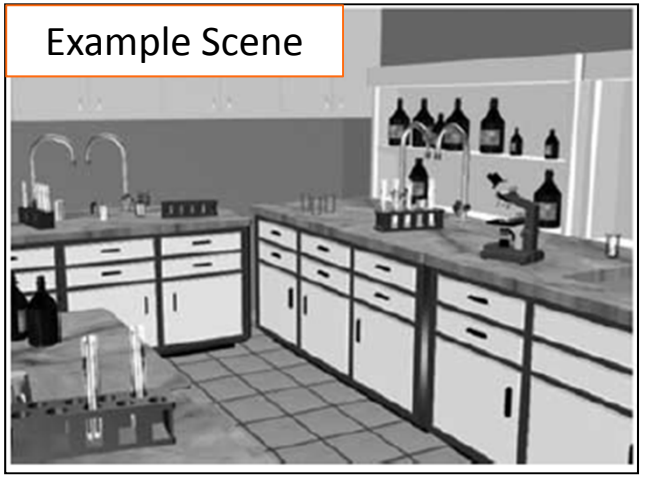


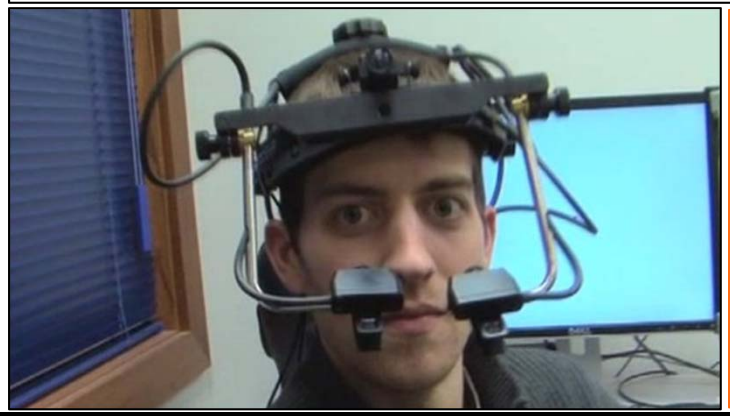
- Dual task: Track red ball with mouse while talking to examine costs of…
- **Speech planning:** current tracking suffers if *next* speech utterance is more complicated
- **Speech production:** current tracking suffers and becomes more variable while producing more complex speech and immediately after

Eye Movements

Fixation duration changes *during* scene viewing based on goals







UNL Psychology Quognitive Program:
Visual Attention, Memory, and Perception Lab

Left: Mark Mills and Eye Tracker

Nested vs. Crossed Multilevel Designs

- When do **items** need to be a separate level-2 **random effect**?
  - Items are clearly nested within persons if the model **fixed effects explain all** of the item variation (no item variation remains)
    - e.g., via item-specific indicators (CFA, IRT)
    - e.g., by item design features given only one item per discrete condition
  - Trials are clearly nested within persons if they are **endogenous**
    - e.g., autobiographical memories, eye movements, speech utterances
  - More ambiguous if trials are **randomly generated** for each person
    - If trials are truly unique per person, then there are no common items…
      … but trials are usually constructed systematically
    - Modeling items/trials as **nested (no item variance) assumes exchangeability**
- When does this matter?
  When turning **experiments** into instruments…

Paradigms in Studying Cognition

Experimental Designs

- Goal is inference about **processes** or **architecture** of cognitive ability
- Create meaningfully different trials through very **specific** manipulations
- Many items given to **few** people
- Multiple aspects of construct represented within a **single task**

- ANOVA → Ability represented by:
  - Mean performance (e.g., RT, # correct)
  - Mean differences between trial conditions

- MLM → Ability represented by:
  - Random intercept
  - Random slopes for trial effects

Psychometric Measures

- Goal is to measure **individual differences** in cognitive ability
- Construct equivalent items to reflect **general** ability being measured
- Fewer items given to **more** people
- **Multiple measures** given to better represent ability construct

- CTT → Ability represented by:
  - Mean performance (e.g., # correct)
  - Mean/component of multiple measures

- CFA/IRT → Ability represented by:
  - Random intercept (≈ factor, theta)
  - Multidimensional ability model

Combining Paradigms

- The **fine-grained task decomposition** found in experimental designs can be combined with latent trait models to more rigorously quantify and predict **individual differences** in ability

- **Synergy** of experimental and individual differences research
  - Theoretical models of cognitive processes inform test construction
  - Research using these instruments then informs theoretical models

- Long-term goal: construct **measures of cognition** that are both theoretically meaningful and psychometrically viable

- Short-term goal: build instruments to better assess individual differences in **selective visual attention**

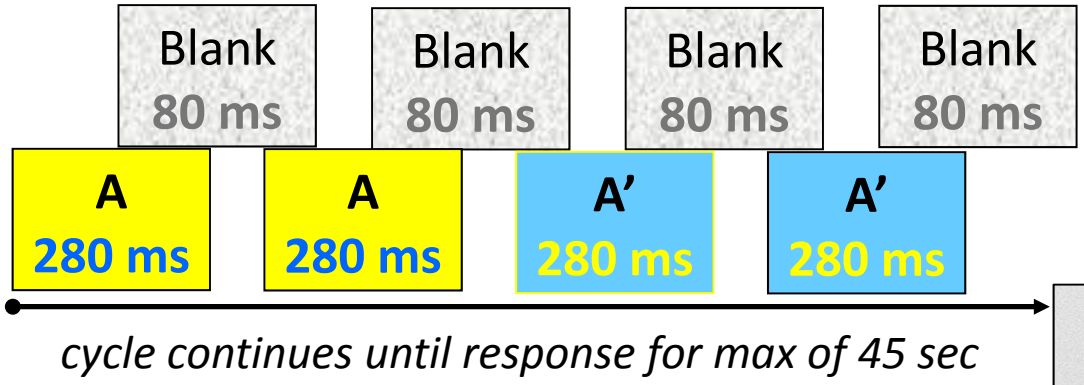
Why Selective Visual Attention?

- Attention is…
  - "A system for routing information and for control of priorities" (Posner, 1980)
  - "The capacity or energy to support cognitive processing" (Plude & Hoyer, 1985)

- Lifespan changes in attentional abilities matter:
  - Significant real-world consequences of attentional deficits with age
  - Difficulty with specific aspects of modulating attention is a marker of some non-normative aging processes

- Measuring visual search:
  - Task difficulty is well-understood → recipe for item creation
  - Current lack of *psychometric* instruments to measure attention
  - Attention is rarely included in individual differences studies, so little is known about how it relates to other abilities (nomothetic span)

Measuring Visual Search Ability: Take 1



cycle continues until response for max of 45 sec

Change detection task using the "flicker paradigm"

Rated Item Design Features:

- Visual clutter of the scene
- Relevance of the change to driving
- Brightness of the change
- Change made to legible sign

- 155 persons, 46 items retained

Measuring Visual Search Ability: Take 1

- How to fit a censored outcome into an "IRT" model?
  - Cut up response time, fit tau-equivalent graded response model
  - "1 immediate" = RT < 8 sec, "2 delayed" = 8-45 sec, "3 time out"

- LLTM version of GRM to examine predictors of item difficulty

$$P\left(Y_{ip} > c \middle| \theta_p\right) = \frac{\exp(\theta_p - \beta_{ic})}{1+\exp(\theta_p - \beta_{ic})}$$

i = item
c = category (threshold)
p = person

- Where each item threshold is:

$$\beta_{ic} = \gamma_{c0} + \gamma_1 Clutter_i + \gamma_2 Relevance_i + \gamma_3 Brightness_i + \gamma_4 Sign_i$$

(difference of category intercepts modeled directly)

- r = .62 of model-predicted and observed item difficulty

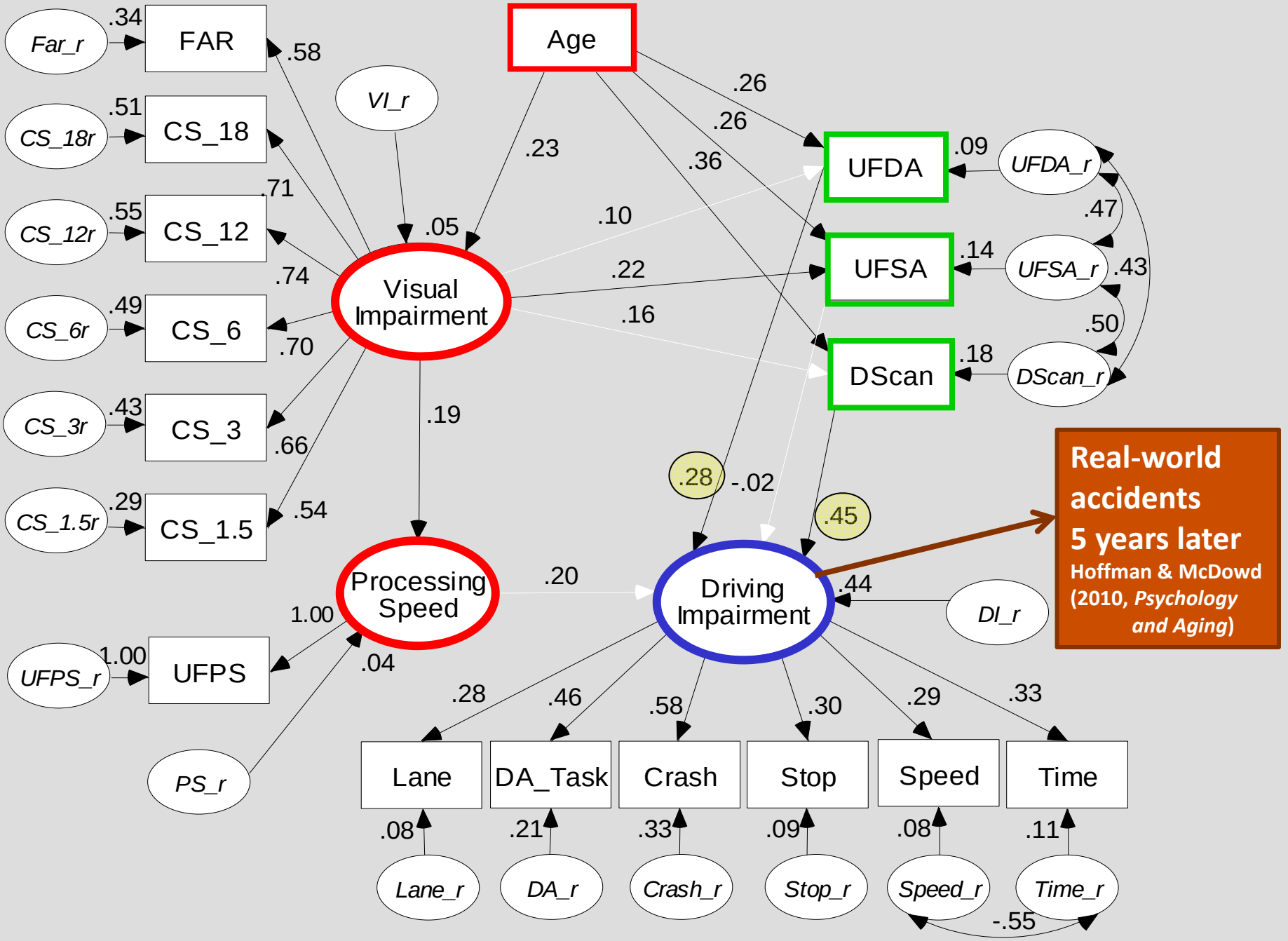
Predicting Driving Impairment*

- 155 current drivers age 63-87; 56% women

- Predictors:
  - Vision (distance acuity, contrast sensitivity)
  - Visual Attention (Useful Field of View subtests, DriverScan)

- Driving Simulator Task Outcome:
  - Easy curves, divided attention, passing, stoplights, obeying speed limits, weaving, narrow radius turns, overtaking vehicles
  - Nothing predicted self-reported and state-recorded accidents

** Like a good neighbor, State Farm was there* (2002 Dissertation Grant)



Hoffman, McDowd, Atchley, & Dubinsky (2005, *Psychology and Aging*)

Take 1: Lessons Learned

- **Response time is problematic as an outcome**
  - Speed contaminated with decision threshold
  - Physical limitations may prevent older adults in responding quickly
  - Continuous, but almost always very skewed distribution
  - Limited utility in real-world assessment

- **Change detection task format is less than ideal**
  - Other-rated item features don't generalize to new items
  - No basis for extrapolation for to create new items
  - Fixed test items can't be used to measure change
  - What if search ability measured was specific to driving scenes?

- **Time for Take 2** → use accuracy and standard search tasks

Measuring Visual Search Ability: Take 2

Project Goals:

1. Determine the **dimensionality of visual attention ability** within and between methods of assessment and its relationships with other constructs (**nomothetic span**)
2. Identify the **factors that predict task difficulty** commonly across both context-free, simple visual search tasks and context-specific, applied visual search tasks measuring selective visual attention (**construct representation**)

UMKC

Joan McDowd

Abilities Measured (by # tasks):

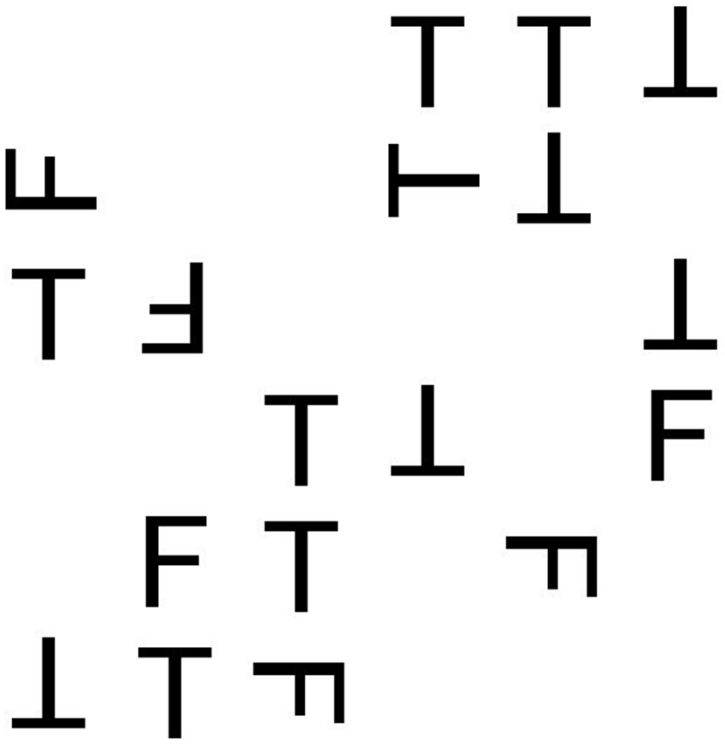
- primary memory (3), working memory (3), comparison speed (3), **visual search (4)**

KU MEDICAL CENTER
The University of Kansas

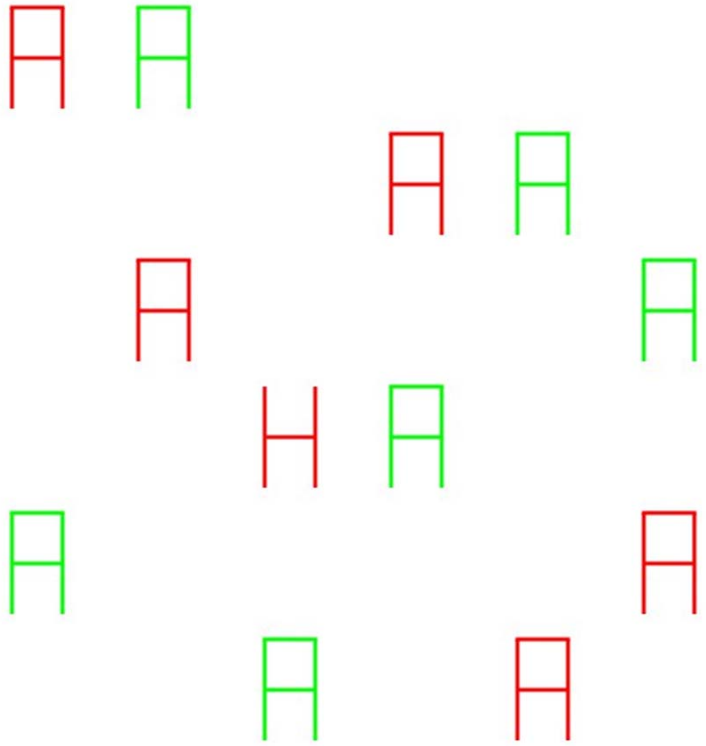
Grayhawk Lab

Context-Free Basic Search Tasks

Rotated-T Search: is top of target T on left or right?



Color-H Search: is target H red or green?

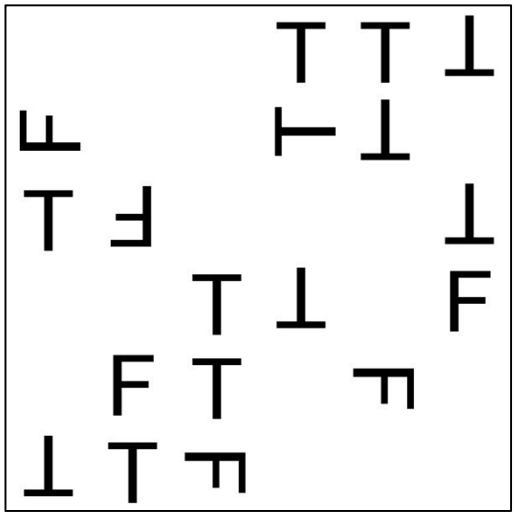


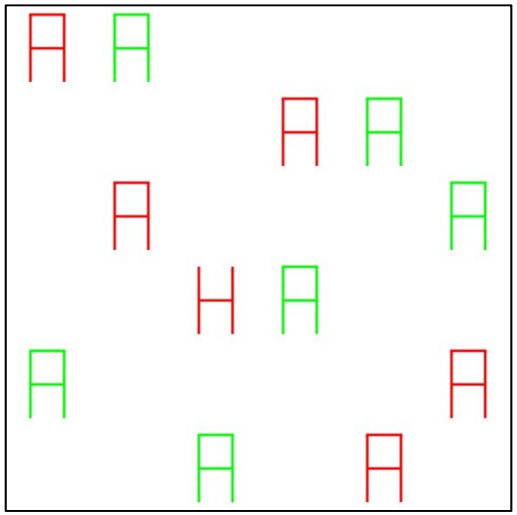
Context-Specific Applied Search Tasks

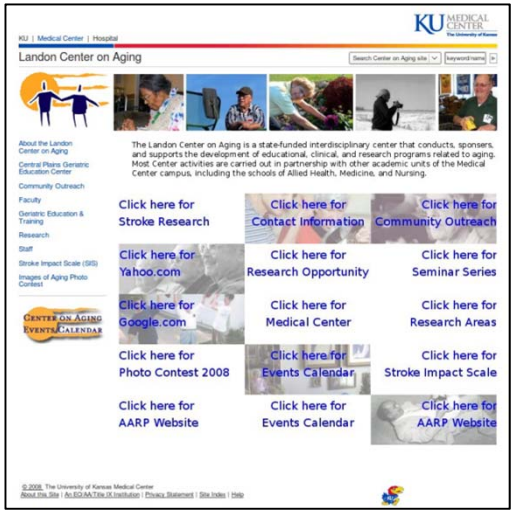
Web Page Search: find link to either “Medical Center” or “Grayhawk Lab”

Grocery Shelf Search: find either can of corn or can of carrots

Measuring Visual Search Ability: Take 2







Each person received 2 of 3 test forms; their order and presentation time were counterbalanced.

Predictors of Accuracy:

- Item presentation time *(short, medium, long)*
- Target location in 6x6 grid *(inner, middle, outer)*
- # distractors *(5 levels)*
- % distractors similar to target *(20, 40, 60, 80, 100)*
- Log of trial order

Sample:

- 329 adults (OA: age 62-88)
- 102 college students (YA)
- Shared medium time, and YA→ short, OA → long

Measuring Ability: IRT Models

- **1PL model** predicts accuracy via fixed item effects and random person effects (i.e., n items are nested in persons)
- **1PL model**:
  - $Probability(y_{ip} = 1|\theta_p) = \frac{\exp(\theta_p - b_i)}{1+\exp(\theta_p - b_i)}$
  - $Logit(y_{ip} = 1|\theta_p) = \theta_p - b_i$

b_i is fixed effect of **difficulty** per item

θ_p is random person ability (variance τ^2_θ)

- **1PL can also be written as generalized multilevel model**:
  - $Logit(y_{ip} = 1|U_{0p}) = \gamma_{00} + \gamma_{10}I_1 + \gamma_{20}I_2 + \cdots + \gamma_{n-1,0}I_{n-1} + U_{0p}$
  - Because item difficulty/easiness is perfectly predicted by the I indicator variables, items do not need a level-2 crossed random effect

$\gamma_{00} + \gamma_{i0}$ is fixed effect of **easiness** per item

U_{0p} is random person ability (variance τ^2_{0P})

Measuring Ability: IRT Models

- 1PL can be extended to **predict item difficulty** via the LLTM
- **LLTM → k** item features predict b_i, random persons (θ_p):
  - $Logit(y_{ip} = 1|\theta_p) = \theta_p - b_i$
  - $b_i = \gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i} + \cdots + \gamma_k X_{ki}$

Item difficulty is predicted via a linear model of X item features and γ fixed effects; **θ_p is random person ability (variance τ^2_θ)**

- **LLTM can also be written as generalized multilevel model**:
  - $Logit(y_{ip} = 1|U_{0p}) = \gamma_{00} + \gamma_{10} X_{1i} + \gamma_{20} X_{2i} + \cdots + \gamma_{k0} X_{ki} + U_{0p}$
  - Because there is no random item effect, the model says that items are still just nested within persons—that item difficulty or easiness is *perfectly* predicted by the X item features (no item differences remain)

Item easiness is predicted via a linear model of X item features and γ fixed effects **U_{0p} is random person ability (variance τ^2_{0P})**

Measuring Ability: IRT Models

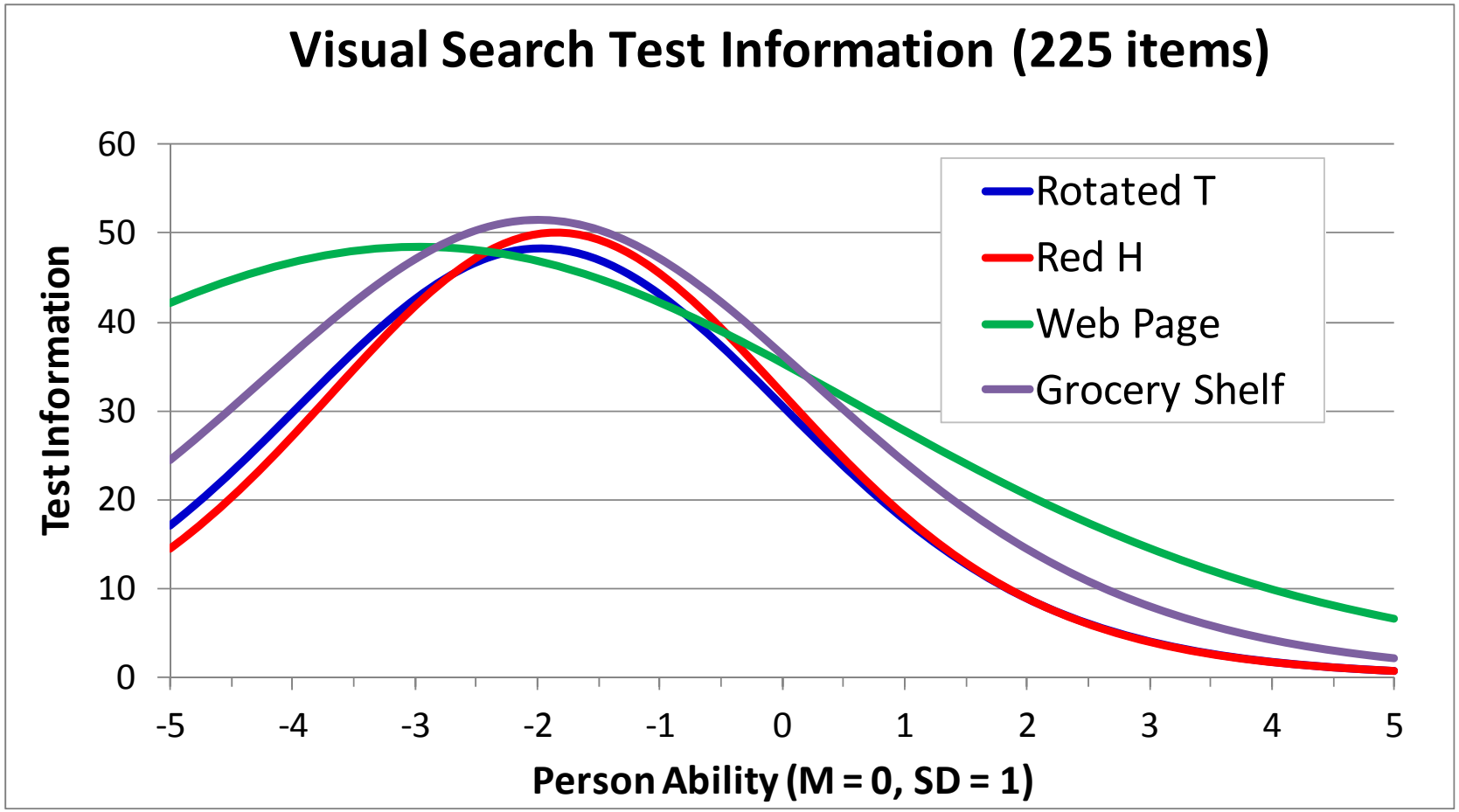
- Experimental tasks can become psychometric instruments via **explanatory IRT (generalized multilevel) models** in which **items** and **persons** have crossed random effects at level 2

$$Logit(y_{tip} = 1 | U_{00p}, U_{0i0}) = \gamma_{00} + \gamma_{10}X_{1ip} + \gamma_{20}X_{2ip} + \cdots + \gamma_{k0}X_{kip} + U_{00p} + U_{0i0}$$

  - U_{0p} is person ability with variance of $\boldsymbol{\tau^2_{0P}}$
  - Item easiness is predicted via a linear model of X item features and γ fixed effects, with random (remaining) variance of $\boldsymbol{\tau^2_{0I}}$
  - Can examine random effects of X item features across persons
  - Can also include person predictors to explain person random effects

- So how did we do? **Reliability for U_{00p}** and **R^2 for τ^2_{0I}**…

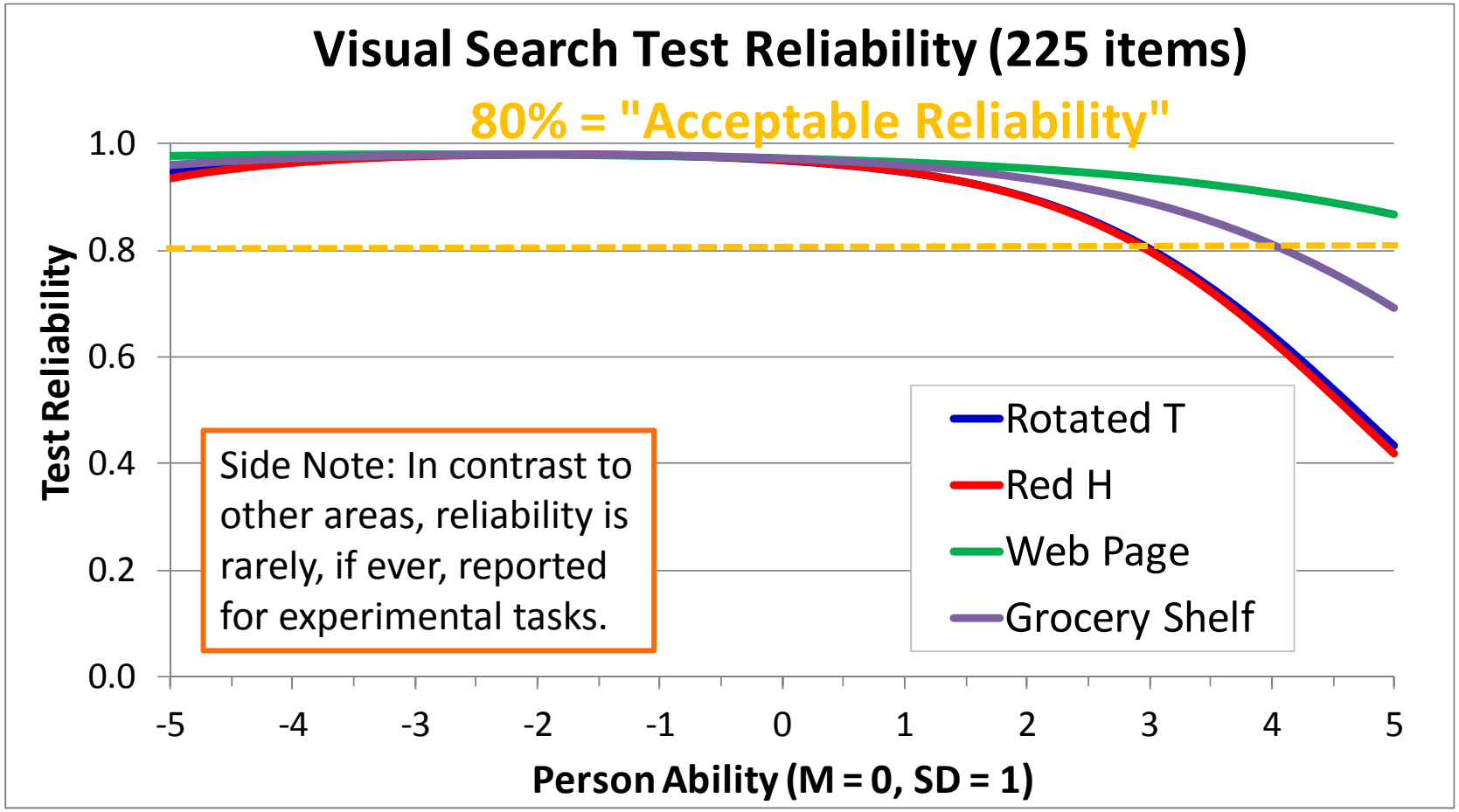
Individual Differences: Test Information



Calculated from model controlling for presentation time only:

$$Logit\left(y_{tip} = 1 \middle| U_{00p}, U_{0i0}\right) = \gamma_{00} + \gamma_{10} Time_{ip} + U_{00p} + U_{0i0}$$

Individual Differences: Reliability



Calculated from model controlling for presentation time only:

$$Logit\left(y_{tip} = 1 \middle| U_{00p}, U_{0i0}\right) = \gamma_{00} + \gamma_{10} Time_{ip} + U_{00p} + U_{0i0}$$

Improving Efficiency (Reducing Boredom)

- Can we give fewer items but still retain measurement precision?
  - ANOVA/CTT: Ability is mean RT or # correct? Then no.
  - MLM/IRT: Ability is estimated along with item properties? Then yes!

- Adaptive search tasks in 5 easy steps:

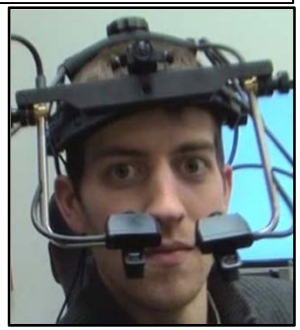
1. **Decompose** item difficulty into effects of known features
2. **Create** new, structurally equivalent items on the fly
3. **Estimate** person ability between each item to determine what level of difficulty the next item should have to be the most informative
4. **Test** younger and older adults via adaptive cognitive tests instead
5. **Change the world of cognition**

THE A - TEAM at UNL

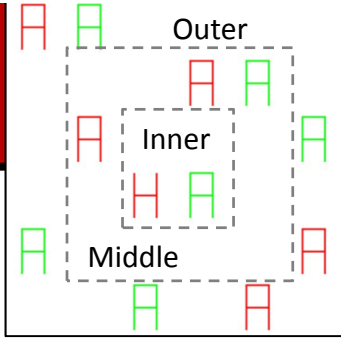
Top: Jonathan Templin (steps 1, 2, 3, 5)

Left: Mark Mills (steps 2, 4, 5)

Right: Lindsey Wiley (steps 4 and 5)

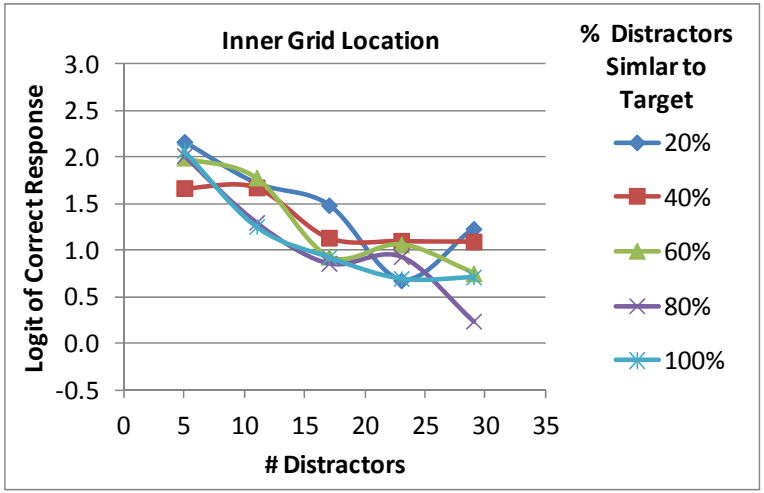


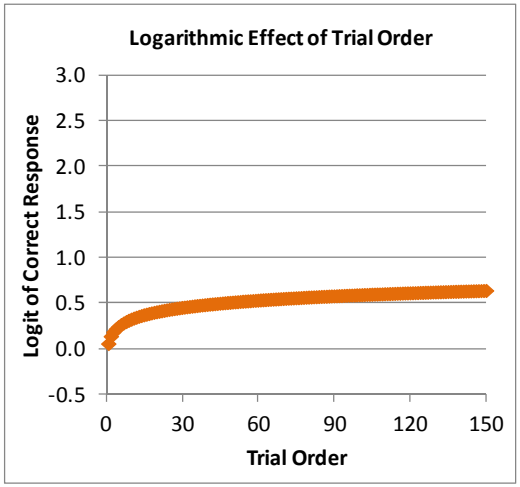
Predicting Random Item Variance



Accuracy of Color-H search: Item R^2 = .84

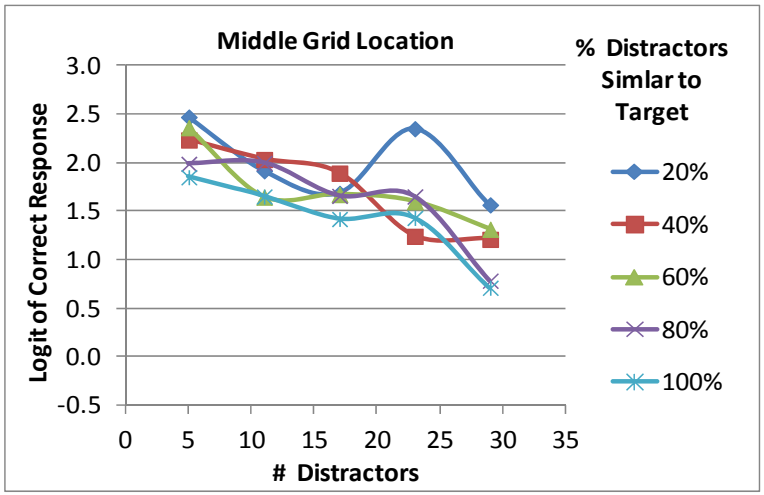
Is a 3-way interaction really *explanatory*? You be the judge…

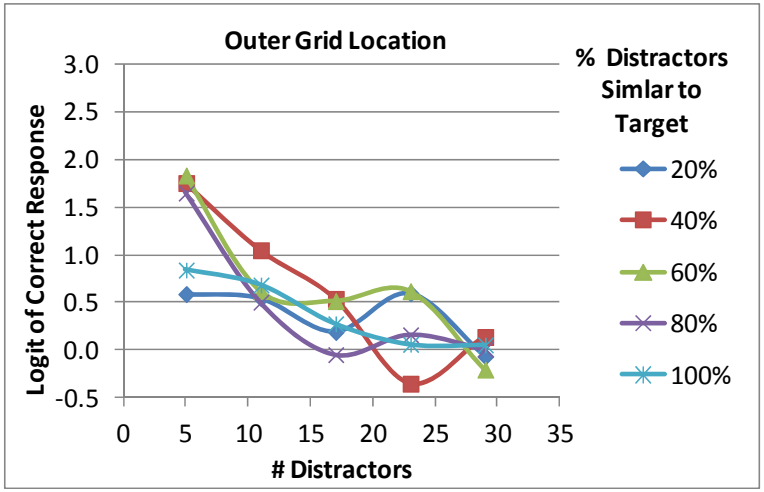




Also included:

- linear effect of presentation time
- time*# distractors, time*grid location
- **83** fixed effects total!





Points to Ponder

- **Issues in on-the-fly item generation in adaptive testing:**
  - "Explained" vs. "described" item variance
  - Invariance of item feature effects by age
  - Using presentation time to fill in the gaps
  - Currently resolving technical difficulties…

- **Future directions:**
  - Converging evidence via multiple estimation approaches
  - Alternative link functions for forced-choice response format
  - Multidimensionality: individual differences in effects of item features via random slopes or latent classes
  - Generalizability: extending approach to other cognitive abilities

Quognitive Psychology: Moving Forward

- **(Non-quognitive-unique) barriers to overcome:**
  - Dissemination of advanced quantitative methods
  - Dogmatism of current practice
  - Multidimensionality of investigators

- **Opportunities for research synergy:**
  - Development of better models for **extreme repeated measures data** (and how to estimate them)
    - e.g., for use with eye-tracking, speech production, ERP, FMRI…
  - A new view on **consultation**:
    Non-quantitative collaborations can be helpful for seeing models used in different contexts, spurring new methodological innovation

Thank you and goodnight!

- Thank you to all who've helped me along the way:
  - Susan Embretson
  - John Flowers
  - Calvin Garbin
  - Scott Hofer
  - Susan Kemper
  - Janet Marquis
  - Joan McDowd
  - Andrea Piccinin
  - Mike Rovine
  - Martin Sliwinski
  - Jonathan Templin
  - Xiangdong Yang
- Comments or questions are welcome: Lesa@unl.edu
- Presentation slides are available for download at: http://psych.unl.edu/hoffman/Sheets/Talks.htm
- Learn more about quantitative psychology at UNL: http://psych.unl.edu/psycrs/index.html